UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/629,175	07/31/2000	Ophir Frieder	7519-164345	4562

26694 7590 09/03/2003

VENABLE, BAETJER, HOWARD AND CIVILETTI, LLP
P.O. BOX 34385
WASHINGTON, DC 20043-9998

EXAMINER	
LE, UYEN T	

ART UNIT	PAPER NUMBER
2171	

DATE MAILED: 09/03/2003

9

Please find below and/or attached an Office communication concerning this application or proceeding.

3

Office Action Summary	Application No.	Applicant(s)	
	09/629,175	FRIEDER ET AL.	
	Examiner	Art Unit	
	Uyen T Le	2171	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *05 June 2003* .

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-43* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-43* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved by the Examiner.

If approved, corrected drawings are required in reply to this Office action.

12)☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____ .

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

a) ☐ The translation of the foreign language provisional application has been received.

15)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ .
4) ☐ Interview Summary (PTO-413) Paper No(s). _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: .

DETAILED ACTION

Response to Amendment

1. The amendment filed 5 June 2003 is objected to under 35 U.S.C. 132 because it introduces new matter into the disclosure. 35 U.S.C. 132 states that no amendment shall introduce new matter into the disclosure of the invention. The added material which is not supported by the original disclosure is as follows: the added feature of "semantic filtering" was not originally claimed and discussed. Figure 2 merely shows parsing a document. To parse by the definition given in Webster's II New Riverside University Dictionary, 1984 at page 856 is to break down into component parts of speech with an analysis of the form, function and syntactical relationship of each part. Thus Figure 2 merely shows syntactic filtering.

Applicant is required to cancel the new matter in the reply to this Office Action.

2. The embedded link at page 2, line 5 is

www.statsci.com/docbrowse/paper/spie98/node1.htm

Applicant is required to remove the www from that line.

3. At pages 1 and 2 of the specification, applicant listed publications that are incorporated by reference. Therefore, the examiner requested amendment to the specification to include the material incorporated by reference. Applicant's statement that no essential subject matter was incorporated by reference is acknowledged. Consequently, objection to the specification is withdrawn.

4. Applicant alleges that support for claim 43 is given in the amendment of December 9, 2002. The examiner disagrees because pages 11-15 of the specification merely discuss syntactic filtering of documents.

5. Applicant seems to try to overcome the Aiken reference of record by adding the limitation of a "single" tuple in claims 1, 29 and "single" hash value in claims 34, 38 and seems to argue the claims as amended. However, the specification provides no support for the claimed "single tuple" or the "single hash value" as now being claimed. Therefore, claims 1-43 are examined as presented in the amendment filed 9 December 2002 and rejection to all claims is maintained using the reference of record, herein repeated.

Claim Rejections - 35 USC § 112

The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

6. Claim 43 is rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the enablement requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. The specification does not provide support for the claimed "semantic filtering" recited at claim 43.

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

7. Claim 43 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention because it is not clear how the "semantic filtering" is performed.

The art rejection of claim 43 is applied as best understood in light of the rejection under 35 U.S.C. 112, first and second paragraphs discussed above.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

8. Claims 1-6, 8, 10-14, 19-21, 23-43 are rejected under 35 U.S.C. 102(a), (e) as being anticipated by Aiken (US 6,240,409).

Regarding claim 1, Aiken discloses a method for detecting similar documents including all the claimed subject matter (see Figures 1a, b, column 3 lines 44-47). The claimed document is met by the string including sub-strings in the method of Aiken (see

column 4, lines 38-53). Note the step of obtaining a document 102, filtering the document 106. Aiken discloses the step of determining a document identifier for the filtered document and a hash value for the filtered document when Aiken shows that the sub-string is associated with a position and a hash value pair (see Figure 2). The claimed step of generating a tuple for the filtered document is met by the fact that a hash value and position pair is created and stored (see step 114, column 6, lines 7-28). Since a string is composed of sub-strings, the hash values computed for sub-strings are clearly computed for documents as claimed. The tuple is clearly compared with a plurality of tuples for the method of Aiken to detect similar strings. Aiken discloses detecting if the document is similar to another document by determining if the tuple is clustered with another tuple in the document storage structured (see Figures 4a, 4b, 4c, column 7, lines 25-34, column 10, line 4- column 12, line 2).

Regarding claim 2, Aiken discloses parsing and filtering the document when Aiken shows removing unimportant words (see column 4, lines 54-67). Clearly the filtered document comprises a token stream of a plurality of tokens as claimed.

Regarding claim 3, Aiken discloses retaining a token according to at least a token threshold when Aiken shows that common or frequent words are removed. A threshold has to be present for the method of Aiken to determine common and frequent words (see column 4, lines 38-53).

Claim 4 merely reads on the translated document in the method of Aiken after all common words and frequent words have been removed.

Regarding claim 5, Aiken discloses determining the hash value for the filtered document by processing individually each retained token in the token stream when Aiken show processing sub-strings and determining their hash values (see column 6, lines 7-28, column 9, lines 24-26).

Regarding claim 6, Aiken discloses determining a score for each token in the token stream and comparing the score for each token to a first token threshold when Aiken shows that common or frequent words are removed. A threshold has to be present for the method of Aiken to determine common and frequent words (see column 4, lines 38-53). The token stream is clearly modified by removing each token having a score not satisfying the first token threshold and retaining each token having a score satisfying the first token threshold as claimed since the common words and frequent words are removed in the method of Aiken.

Regarding claim 8, Aiken discloses filtering by removing from the token stream at least one token corresponding to a stop word (see column 4, lines 57-58, column 8, line 67- column 9, line 3).

Regarding claim 10, Aiken discloses removing a token from a token stream based on collection statistics and at least one token threshold when Aiken shows that the method remove words of "the" "and" , "this", "is" (see column 4, lines 57-58, column 8, line 67- column 9, line 3). A token threshold has to be present for the method of Aiken to determine common and frequent words (see column 4, lines 38-53).

Regarding claim 11, Aiken discloses removing a token from a token stream (see column 4, lines 57-58, column 8, line 67- column 9, line 3).

Regarding claim 12, Aiken discloses removing formatting from the document (see column 4, lines 55-57).

Regarding claims 13, 14, clearly the method of Aiken uses collection statistics pertaining to a plurality of documents for filtering the document since the input file is compared to a set of collected files to detect similarity (see column 2, lines 47-51). The collection statistics have to be present for the collected documents to be clustered as shown in the method of Aiken (see Figure 4c, column 11, line 47- column 12, line 2).

Regarding claim 19, Aiken discloses a hash table (see column 12, lines 40-44).

Regarding claim 20, Aiken discloses that the document storage structure comprises a tree (see column 8, lines 30-38).

Regarding claim 21, Aiken discloses that the tree comprises a binary tree (see column 8, lines 36-38).

Regarding claim 23, Aiken discloses a hash table and at least one tree (see column 5, lines 33-40, column 8, lines 30-38).

Regarding claim 24, Aiken discloses inserting the tuple into the document storage structure (see Figure 1a, 1b, 4a, 4b, 4c).

Regarding claim 25, the hash table of Aiken clearly comprises a plurality of bins of tuples as claimed and the step of determining if the tuple is clustered with another tuple clearly comprise determining if the tuple is co-located with another tuple at a bin of a hash table (see Figures 1, 2, 4c, column 7, line 46- column 8, line 33).

Regarding claim 26, Aiken discloses a tree comprising a plurality of branches, each bucket of the tree comprising at least one tuple and wherein the step of

determining if the tuple is clustered with another tuple clearly comprise determining if the tuple is co-located with another tuple in a bucket of the tree (see column 8, lines 31-54, Figure 4c).

Claims 27, 29 correspond to a system to perform the method of claim 1, thus are rejected for the same reasons stated in claim 1 above.

Claim 28 corresponds to a computer program product to perform the method of claim 1, thus is rejected for the same reasons stated in claim 1 above.

Claim 30 is a mere combination of claims 1-4, 26, thus is rejected for the same reasons stated in claims 1-4, 26 above.

Claims 31, 33 correspond to a system to perform the method of claim 30, thus are rejected for the same reasons stated in claim 30 above.

Claim 32 corresponds to a computer program product to perform the method of claim 30, thus is rejected for the same reasons stated in claim 30 above.

Regarding claim 34, Aiken discloses all the claimed subject matter including determining a hash value for a document (see Figure 1, column 4, line 17- column 7, line 45, column 9, lines 16-30), accessing a document storage structure comprising a plurality of hash values, each hash value representing one of a plurality of documents (see Figure 4a, column 10, line 4- column 11, line 46), determining if the hash value is equivalent to another hash value in the document storage structure (see Figure 4c, column 11, line 47- column 12, line 2). Note the document is met by the strings including sub-strings in the method of Aiken (see column 4, lines 38-53).

Claims 35, 37 correspond to a system to perform the method of claim 30, thus are rejected for the same reasons stated in claim 34 above.

Claim 36 corresponds to a computer program product to perform the method of claim 30, thus is rejected for the same reasons stated in claim 34 above.

Regarding claim 38, Aiken discloses a method for detecting similar documents including comparing a document to a plurality of documents in a document collection using a hash algorithm and collection statistics to detect if the document is similar to any of the documents in the document collection (see Figures 1a, b, 4a, 4b, 4c). The claimed collection statistics have to be present and used in the method of Aiken for the similar documents to be clustered (see Figure 4c, column 11, line 47- column 12, line 2).

Regarding claim 39, clearly the collection statistics pertain to the document collection since the statistics are used in clustering the collected documents (see Figure 4c).

Claims 40, 42 correspond to a system to perform the method of claim 38, thus are rejected for the same reasons stated in claim 38 above.

Claim 41 corresponds to a computer program product to perform the method of claim 38, thus is rejected for the same reasons stated in claim 38 above.

Regarding claim 43, Aiken discloses the step of performing semantic filtering on the document when Aiken shows that each string is translated according to rules tailored to the type of document being translated such as the syntax and semantic of a particular programming language (see column 4, lines 47-53).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

9. Claims 7, 9, 15-18, 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Aiken (US 6,240,409).

Regarding claim 7, although Aiken does not specifically show the step of comparing the score for each retained token to a second token threshold and modifying the token stream as claimed, Aiken explicitly show that not every substring's hash value is stored (see column 6, lines 29-30). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed feature while implementing the method taught by Aiken in order to further filter the document and save memory.

Regarding claim 9, although Aiken does not explicitly disclose filtering by removing a duplicate of another token in the token stream, it would have been obvious to one of ordinary skill in the art to include such a feature in order to avoid processing redundant token, thus saving time and resources.

Regarding claims 15-18, although Aiken does not explicitly show that the method uses specific hash algorithms as claimed, it is notoriously well known in the art to use different hash algorithms depending on users' requirements. Therefore, it would have been obvious to one of ordinary skill in the art to include all the claimed features while implementing the method of Aiken in order to suit users' needs.

Regarding claim 22, although Aiken does not explicitly show that the binary tree is balanced, it would have been obvious to one of ordinary skill in the art to include such a feature while implementing the method of Aiken in order to store data efficiently and to facilitate searching and localization.

Conclusion

10. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Meyerzon et al (US 6,547,829) teach detecting duplicate documents in web crawls.

Sowa et al (US 6,594,665) teach storing hashed values of data in media to allow faster searches and comparison of data.

11. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Uyen T Le whose telephone number is 703-305-4134. The examiner can normally be reached on M-F 7:00-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Safet Metjahic can be reached on 703-308-1436. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

Uyen Le
Primary Examiner
AU 2171

20 August 2003